DOI:
10.1016/j.neuroimage.2014.01.034

Document Version
Preprint – the version submitted for publication

Citation for published version (APA):
Wright, R., Kyriakopoulou, V., Ledig, C., Rutherford, M., Hajnal, J., Rueckert, D., & Aljabar, P. (2014). Automatic quantification of normal cortical folding patterns from fetal brain MRI. NeuroImage, 91, 21-32. 10.1016/j.neuroimage.2014.01.034

Automatic Quantification of Normal Cortical Folding Patterns from Fetal Brain MRI

R. Wright[a,*], V. Kyriakopoulou[b], C. Ledig[a], M.A. Rutherford[b], J.V. Hajnal[b], D. Rueckert[a], P. Aljabar[a,b]

[a] *Biomedical Image Analysis Group, Department of Computing, Imperial College London, SW7 2AZ, UK*

[b] *Imaging Sciences & Biomedical Engineering Division, King's College London, St. Thomas' Hospital, SE1 7EH, UK*

Abstract

We automatically quantify patterns of normal cortical folding in the developing fetus from *in utero* MR images (N=80) over a wide gestational age (GA) range (21.7 to 38.9 weeks). This work on data from healthy subjects represents a first step towards characterising abnormal folding that may be related to pathology, facilitating earlier diagnosis and intervention. The cortical boundary was delineated by automatically segmenting the brain MR image into a number of key structures. This utilised a spatio-temporal atlas as tissue priors in an expectation-maximization approach with second order Markov random field (MRF) regularization to improve the accuracy of the cortical boundary estimate. An implicit high resolution surface was then used to compute cortical folding measures. We validated the automated segmentations with manual delineations and the average surface discrepancy was of the order of 1mm. Eight curvature-based folding measures were computed for each fetal cortex and used to give summary shape descriptors. These were strongly correlated with GA ($R^2 = 0.99$) confirming the close link between neurological development and cortical convolution. This allowed an age-dependent non-linear model to be accurately fitted to the folding measures. The model supports visual observations that, after a slow initial start, cortical folding increases rapidly between 25 and 30 weeks and subsequently slows near birth. The model allows the accurate prediction of fetal age from an observed folding measure with a smaller error where growth is fastest. We also analysed regional patterns in folding by parcellating each fetal cortex using a nine-region anatomical atlas and found that Gompertz models fitted the change in lobar regions. Regional differences in

*Corresponding author
Email address: r.wright11@imperial.ac.uk (R. Wright)

growth rate were detected, with the parietal and posterior temporal lobe exhibiting the fastest growth, while the cingulate, frontal and medial temporal lobe developed more slowly.

Keywords: fetal MRI, cortical folding, brain development, Gompertz function

---

1. Introduction

In utero magnetic resonance imaging (MRI) has emerged as an important clinical analysis and diagnostic tool for prenatal abnormality in recent years (Hosny and Elghawabi, 2010). With the development of ultrafast scanning techniques such as single-shot fast spin-echo (SSFSE), detailed 2D image slices, free from motion artifact, can be acquired without invasive fetal curarization and maternal sedation (Glastonbury and Kennedy, 2002). Ultrasound is inexpensive and widely available and thus remains the modality of choice for initial screening. With significantly greater tissue contrast and spatial resolution, MRI affords an opportunity for more detailed investigation of abnormalities, thus enabling more accurate diagnoses for subjects referred after screening.

The advent of slice-to-volume reconstruction techniques (Rousseau et al., 2006; Jiang et al., 2007) revealed coherent 3D MR images of the fetal brain for the first time with a high signal-to-noise ratio (SNR) (Figure 1). Super-resolution techniques (Rousseau et al., 2010; Kim et al., 2010; Gholipour et al., 2010; Kuklisova-Murgasova et al., 2012) have subsequently boosted spatial resolution by exploiting repeated overlapping slice samples. These techniques have recently stimulated new research into the structural development of the fetal brain in the neuroimaging community (Gholipour et al., 2012; Habas et al., 2012; Jacob et al., 2011; Rajagopalan et al., 2011; Scott et al., 2011; Serag et al., 2012; Caldairou et al., 2011; Habas et al., 2010; Scott et al., 2013).

Cortical folding, or gyrification, is perhaps the most striking structural change that occurs in the fetal brain during gestation. The neocortex rapidly becomes more convoluted from approximately 25 weeks, leading to a dramatic increase in surface area without an equivalent increase in thickness. This allows a greater number of neurons to be packed into the limited intra-cranial space and is thought to be critical to many human advances in intelligence (Lui et al., 2011). SSFSE imaging produces strong contrast between cortical grey matter (GM) and adjacent white matter (WM) allowing excellent visualisation of this process *in utero*. Depending on the severity, malformations of the cortex can cause a range of outcomes including death in infancy, psychomotor retardation, developmental delay and seizures (Ghai et al., 2006). Lissencephaly (smooth brain) is a common result of abnormal neuronal migration where folding is either reduced or completely absent. Prognosis is usually poor and related to the degree

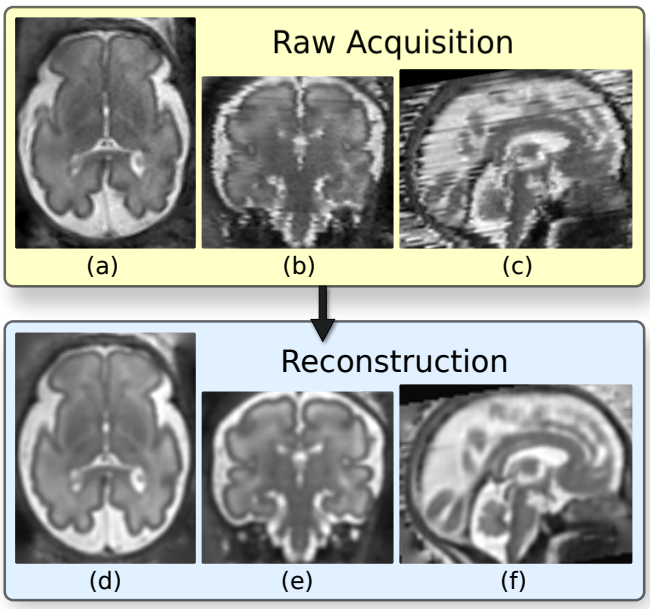


Figure 1: **Reconstruction of 3D volume from 2D slices.** 2D slices are acquired in the transverse plane using a fast scanning sequence to freeze motion (a). The coronal (b) and sagittal (c) views are corrupted by motion between slice acquisitions. The motion between slices is estimated and a 3D volume is reconstructed from the original slice data providing a coherent 3D image (d,e,f).

of smoothness, but early diagnosis could allow better care for the patient. Cortical folding is critical to normal development and is well visualised and therefore is an obvious focus for a population based study.

Segmenting the brain into different anatomical structures is an important first step towards the analysis of developmental processes. Manual segmentation is slow, expensive, and requires expert knowledge. Therefore, a robust automatic segmentation algorithm is highly desirable for large population based studies. However, automated segmentation of the fetus is particularly challenging for a number of reasons and techniques developed for the adult brain tend not to perform well. Firstly, rapid brain development during gestation creates large anatomical variations within fetal cohorts (Figure 2). The size and shape of structures can change significantly in a matter of weeks and an observable structure in an older fetus may not be observable in a younger fetus. Furthermore, tissue types of the fetal brain have relatively heterogeneous and overlapping signal intensities due to ongoing developmental processes at the cellular level, such as WM myelination. The imaging may also be limited by some of the necessary compromises that are made to minimise acquisition time and freeze fetal motion. To retain acceptable SNR with a short acquisition sequence, thick slices of

around 3mm are typically acquired (Prayer et al., 2004). This leads to partial volume (PV) voxels where adjacent tissues meet, resulting in an intermediate intensity signal. For example, in T2 weighted SSFSE images at the cerebral boundary, low signal from GM combines with high signal from cerebrospinal fluid (CSF), resulting in a moderate intensity similar to WM. This is more apparent in the deep narrow sulci of older fetuses and segmentation algorithms that rely on intensity information can incorrectly label PV voxels as WM in these areas. With adult structural MRI, T1-weighted relaxation is normally used to give better tissue contrast, where WM has the largest intensity followed by GM then CSF. In this instance a PV voxel at the cortical boundary will not replicate the signal response of another tissue, thus mislabelling of PV voxels is a lesser confound. In fetal scanning, the problem has also been compounded by precautionary limits to magnetic field strength. Neuroscientific studies predominately utilise 3 Tesla (T) scanning devices, however 1.5T is typically used for fetal MRI, which limits SNR for a given spatial resolution and acquisition time (Welsh et al., 2011). MRI is considered a relatively safe, non-invasive modality, however there has been concerns previously over the effect of noise and heat on the fetus when using stronger magnetic fields (Wilde et al., 2005). No formal studies exist on this subject to date, although numerical simulations of specific absorption rate have predicted that average fetal temperature would be within international safety limits (Hand et al., 2010) and studies of the fetal brain using 3T scanners are now beginning to emerge (Egañ-Ugrinovic et al., In Press).

To date, few methods have been developed specifically to segment the fetus from *in utero* MRI. A popular segmentation approach for adult brain images, expectation-maximization segmentation with Markov random field regularization (EM-MRF) (Van Leemput et al., 1999) has been used for segmenting the fetus (Habas et al., 2010). EM segmentation requires a probabilistic tissue atlas to give prior spatial knowledge of tissue likelihood. In this work we address the large variation in anatomy by employing a spatio-temporal atlas (Serag et al., 2012), where an age-specific probabilistic atlas is used to capture the anatomy at several time-points during development. This spatio-temporal atlas technique has also been used to segment the developing brain in young children and neonates (Serag et al., 2012; Murgasova et al., 2007; Kuklisova-Murgasova et al., 2011). This segmentation framework has been the basis for a number of normal (Habas et al., 2012; Rajagopalan et al., 2011; Scott et al., 2011) and abnormal (Scott et al., 2013) brain development studies in fetuses ranging from 20 to 30 weeks gestational age. EM-MRF is sufficient to segment fetuses around this gestational age range, however it fails to explicitly tackle the problem of mislabelled PV voxels at the cortical boundary and therefore it is ill-suited to segmenting older fetal brains. Incorporating constraints to reduce mislabelled PV voxels into an EM framework has been shown to reduce this problem at the cortical boundary in both neonates and fetuses (Xue et al., 2007; Ledig et al., 2012; Wright et al., 2012).

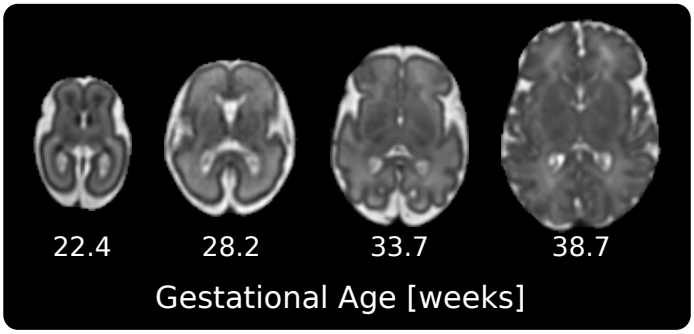


Figure 2: **Diversity of anatomy.** These MR images illustrate the large variation in brain morphology during fetal development.

Several studies have attempted to segment the fetal cortex and characterise its development (Rajagopalan et al., 2011; Habas et al., 2012; Scott et al., 2013). These have been limited to a small number of subjects and have focused on younger fetuses, which are easier to segment due to simple anatomy. The most extensive study of normal folding patterns to date concentrated on folding patterns and asymmetries of 40 fetuses from 20 to 28 weeks gestational age (Habas et al., 2012). Our objective is to quantify normal folding patterns with a much larger cohort from 22 weeks up to term. Our goal is to extend the understanding of how the cortex folds through this critical developmental period. In this work, we present a complete image analysis framework for quantifying fetal folding patterns from *in utero* MRI (see Figure 3 for an overview). We apply a spatio-temporal model of average MR intensities and tissue probabilities to automatically segment 80 normal fetal controls, over a wider gestational age range than has previously been studied (22-39 weeks), into a number of key tissue types using an EM based segmentation framework as described in (Ledig et al., 2012). This framework incorporates a regularization term to specifically penalise mislabelled PV voxels around the cortical boundary in order to improve segmentation accuracy. Sixteen of these subjects were segmented manually to validate the accuracy of the automated procedure. An implicit cortical surface representation was formed on a high resolution voxel grid and principal curvatures were evaluated over the surface voxels using the structure tensor and Hessian matrix. Three curvature-based shape descriptors were computed at each surface voxel. Eight folding measures were then computed for each fetus, each of which summarised a shape descriptor over the surface to produce a single scalar. These folding measures were used to quantify the degree of gyrification through this period of development. To analyse the local patterns of sulcal development the delineated cortical surface was parcellated into nine regions for each fetus by registering it to an anatomical atlas. Folding measures were then computed for each region individually and analysed in relation to development.

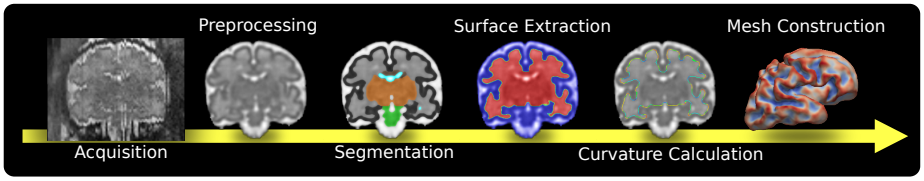


Figure 3: **Processing pipeline overview**. A raw acquisition consisting of multiple overlapping slices is reconstructed to a complete 3D volume free from motion artifacts. The intra-cranial region is then automatically extracted. The brain is automatically segmented into 8 structural regions using an expectation-maximization framework with Markov random field regularization. A surface is then defined by combining the posterior probability maps of structures within the inner cortical boundary. Principal curvatures were calculated at each boundary location and global and regional summary measures were derived from them to quantify folding patterns. Curvature descriptors were visualized on a mesh, computed using the marching cubes algorithm (Lorensen and Cline, 1987).

2. Material and Methods

2.1. Data and Pre-processing

A cohort of 80 clinically certified healthy fetal controls was used for the study with a gestational age range of 21.7 to 38.9 weeks (mean 29.6±4.6) (Figure 4). T2-weighted MR imaging was performed on a 1.5T Philips Intera scanner with a 32 channel coil. An SSFSE sequence was used to obtain individual 2D slices that were free from motion artifacts. Multiple overlapping slices were acquired over the region of interest and then reconstructed using a robust slice-to-volume registration method (Jiang et al., 2007), producing a volume with an isotropic voxel resolution of approximately $1mm$. Images were processed using N4 (Tustison et al., 2010) to correct for intensity non-uniformity to achieve a consistent, spatially invariant, signal intensity distribution for each tissue.

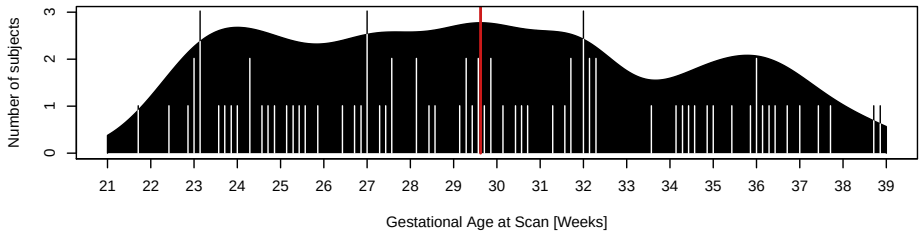


Figure 4: **Cohort age spread.** Bar chart with accompanying density trace, estimated using a Gaussian kernel ($\sigma = 1$). The mean age of the cohort is shown in red.

2.2. Automatic Brain Extraction

Before segmenting the brain into different structures, it is useful to define a region of interest bounded by the skull, discarding any MR signal outside the region. With adult subjects, the strong signal contrast between the skull and intra-cranial matter can facilitate automatic extraction of the intra-cranial region. Popular extraction algorithms such as brain extraction tool (BET) (Smith, 2002) exploit this, requiring little prior knowledge to accurately extract brain matter. With fetal subjects, however, this task is significantly more challenging as the boundary of the skull is less prominent due to heterogeneous signal from stationary uterine tissue and amniotic fluid immediately surrounding the head of the fetus. As a result, methods developed for adults that rely on a strong intensity boundary are ill-suited and perform poorly when applied to fetal scans.

We applied a patch-based approach to this problem (Eskildsen et al., 2012), which we found to extract fetal brains accurately. This approach relies on the assumption that a similar patch of voxels in an input image and atlas image are likely to be centred on the same tissue. This gives a mechanism for non-local, many-to-one, label propagation. By aligning an input image with an atlas, the search for similar patches can be restricted to a local neighbourhood, which is likely to be sufficient, given the limited variability in both shape and location of the skull. Numerous similar atlas patches form a labelling consensus for each voxel in an input image, with each atlas label weighted by the similarity of its surrounding patch and input image patch, ensuring the most similar patches are the most influential (Figure 5). For a more detailed description of this procedure the reader is referred to Appendix A.

As the accuracy of the atlas set is critical for accurate extraction, we constructed fetal atlases by combining manual delineations of the cerebrum, brain-stem, lateral ventricles and extra-cerebral CSF, filling the small remaining voids (cavum, third ventricle) automatically. A set of 35 atlases was generated, each consisting of an MR image and a binary mask with foreground voxels defining the matter inside the skull. We then mirrored each atlas along the centre-line of the brain, effectively doubling the size of the atlas set to 70 in total.

Before extraction, the atlas set and input image were all co-aligned and intensity normalized. By aligning the images, and only considering patches in a small local neighbourhood, the computational cost is reduced considerably and the influence of dissimilar distant patches is removed. For this purpose, all atlases were affinely registered to a target subject of mean gestational age within the cohort and each new input image was then registered to this target before labelling. Patch similarity was evaluated using the sum of squared difference (SSD), therefore all the images were normalized using histogram matching (Nyúl and Udupa, 1999) to standardise the intensity scale, allowing SSD to be used as a meaningful similarity metric.

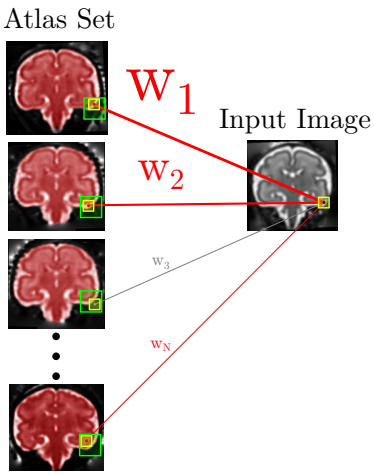
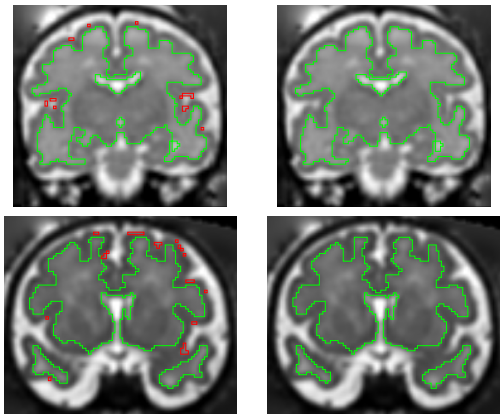


Figure 5: **Patch search and weighting.** The input image and atlas images are co-aligned so that similar patches (yellow) can be found within a local neighbourhood (green) around an input voxel. The patch similarity can then give a weighting when fusing the atlas labels. The reader is referred to Appendix A for a detailed explanation of the label fusion process.

To assess the accuracy of the brain extraction, a leave-one-out cross-validation was performed which resulted in a high mean Dice's coefficient of 0.983 using 24 atlases, a patch diameter of 7 voxels and a neighbourhood diameter of 5 voxels.

2.3. Automatic Segmentation

The discrete nature of MR images results in PV voxels that contain an intermediate signal response at locations where adjacent tissues meet. At the cortical boundary in T2-weighted images, PV voxels with low signal response from cortical GM and high signal response from CSF have a moderate intensity indistinguishable from WM, and are frequently mislabelled. This effect is pronounced in fetal MRI because of the slice thickness necessary to achieve an acceptable SNR when restricted to a short acquisition time (to freeze motion) and a relatively low field strength (to protect the fetus). To address this problem we use an EM segmentation framework that incorporates a standard MRF as well as additional second order MRF penalisation term, which was evaluated using neonatal MR images (Ledig et al., 2012), where a similar PV problem exists. This second order MRF can penalise a specific tissue class based on its adjacency with two other specific classes, as opposed to just one in the standard MRF formulation. By penalising WM when it is implausibly adjacent to both cortex and CSF we were able to improve segmentation accuracy at the cortical boundary (Figure 6).



(a) (b)

Figure 6: **Improved cortical segmentation using a second order MRF**. Two subjects are shown, each segmented using a standard EM-MRF algorithm (a) and an EM algorithm incorporating a second order MRF (b). Voxels correctly labelled as WM are shown in green while PV voxels mislabelled as WM are shown in red. Note the reduction in mislabelled voxels in (b).

MRI volumes were automatically segmented into a background class and 7 foreground classes: cortical GM, deep grey nuclei, germinal matrix, WM, non-cerebral tissue (brain stem and cerebellum), intra-cerebral CSF, and extra-cerebral CSF. For this, a non-rigid spatio-temporal atlas of average tissue intensities, represented by a sequence of MR templates (Serag et al., 2012) for each week of gestation with associated tissue probabilities (Wright et al., 2012), was used as prior information to guide the segmentation process. In this previous work, the atlas at each age was constructed by non-rigidly registering all contributing images in a pairwise manner using the free-form deformation (FFD) framework (Rueckert et al., 1999), and, for each image in turn, averaging the resulting transformations to the other images, giving a mapping towards a mean MR template. Templates and associated tissue probability maps were created by averaging the transformed image intensities and segmentations. Subject specific tissue probability maps were constructed on a leave-one-out basis to remove bias from manual segmentations, i.e. manual segmentations for each subject did not contribute to its own prior tissue probabilities. For further details of the atlas construction, the reader is referred to (Serag et al., 2012). Before segmentation, the atlas template with the closest age to the subject was non-rigidly registered to the input image. The associated tissue probability maps were subsequently transformed onto the native voxel grid of the subject before segmentation.

The FFD control point spacing used for pairwise registration during atlas construction controls the flexibility of the local deformation between subjects and has a considerable impact on the sharpness of the average MR intensity tem-

plate and its associated tissue probability maps. As the control point spacing is reduced, the registration becomes less robust and inherently more susceptible to misalignment errors. As a result, the registration algorithm may converge to local minima. Furthermore, as atlas priors become sharper, they have stronger influence over a resulting segmentation. In this case, an EM algorithm will be dominated by the prior information and fail to compensate for registration misalignments, leading to a segmentation that does not appropriately describe the subject's anatomy. Conversely, with a coarse control point spacing, the MR template and probability maps will be considerably less sharp. In this case, when an atlas prior is aligned with an input image, there can be significant WM probability overlapping cortical sulci and vice versa. This can lead to an increase in partial volume misclassification (Serag et al., 2012).

Therefore, we have used a relatively fine control point spacing of 1.5mm for registration during atlas construction, which produces crisp tissue priors that we can then relax (e.g. by blurring) during segmentation to compensate for misalignment of the original tissue priors. This relaxation step occurs after the EM algorithm has first converged to a solution, before continuing.

An FFD spacing of 3mm was chosen for registering an age-matched template to each input image to reflect the flexibility used in atlas construction, while giving a greater tolerance for registration error. This is desirable as during atlas construction, inter-subject registrations are averaged, reducing the prominence of segmentation errors.

After segmentation, a connected-component analysis was used to restrict the sub-cortical matter to a single mass. As the cortex becomes more convoluted, considerable PV exists in narrow sulci. By constraining sub-cortical matter to a single mass, we effectively prohibited the probability of WM in these regions. Posterior probability maps for sub-cortical regions (germinal matrix, WM, deep grey nuclei) and intra-cerebral CSF were combined and then binarized using a threshold of 0.95. The largest connected component (LCC) was then determined leaving a collection of residual components. The binarization threshold was chosen heuristically as the highest value which did not result in WM voxels inside thin gyri becoming disconnected from the LCC. Residual component voxels and any remaining voxels with a non-zero probability that were closer to a residual component than the largest connected component were assumed to be sulcal PV voxels and the corresponding probability was reassigned as external CSF in the probability map. This was repeated iteratively using a descending threshold to remove all mislabelled WM.

2.4. Voxel-based Surface Representation

The surface of the brain is not an ideal location for measuring curvatures. As gyrification proceeds, sulci narrow, adjacent gyri begin to squeeze against each

other, and it becomes increasingly difficult to delineate the cerebrum boundary. Furthermore, inexact brain extraction may remove small amounts of cortical GM where the cerebrum is very close to the skull. The inner cortical boundary, however, is easier to delineate and robustly segment, therefore we chose this boundary on which to compute curvatures.

The surface was approximated directly on an image's voxel grid by combining posteriors for 4 classes, germinal matrix, WM, the deep grey nuclei and intracerebral CSF, forming a probability map for matter lying inside of the cerebral cortex. This map was smoothed using a Gaussian kernel with a full width at half maximum of 2mm which was found to ameliorate segmentation irregularities arising from imperfections in the reconstruction process while retaining the fidelity of the estimated boundary. The map was then binarized using a threshold that preserved its volume and the largest connected component was retained. A surface voxel was then defined as any foreground voxel that shared a face with a background voxel.

A 1mm voxel grid is too coarse to give an accurate surface representation using a binary volume and sub-voxel accuracy is required to capture the curvature of the convoluted cortex. Therefore, the input image was upsampled by a factor of 3 and a second segmentation step was performed using the previously calculated posteriors as priors. This resolution was found to be sufficient for accurate surface representation while retaining acceptable computation time and memory requirements. Further upsampling was also tested, but there was little visible difference between surfaces when visualised as a mesh

2.5. Folding Measures

At each of the surface voxels, the principal curvatures k_1 and k_2 were calculated from the structure tensor and Hessian matrix using the techniques described in (Rieger et al., 2004), which have previously been applied to neonates (Rodriguez-Carranza et al., 2008). The interested reader is referred to Appendix B for more details. The curvature of a curve at a surface point p is defined as the reciprocal of the radius of a circle passing through p and a pair of additional points on the curve infinitesimally close to p. The curvature can be measured along any plane intersecting the surface which contains the surface normal. The principal curvatures k_1 and k_2 are the maximum and minimum curvature values at point p. These scalars are fundamental to a number of local shape descriptors that we will use in our curvature analysis, including mean curvature, Gaussian curvature and curvedness (Table 1)).

The principal curvatures and the derived descriptors are dependant on the scale of the surface and do not reflect its complexity. For example, the curvature of a sphere varies depending on its size. To remove the influence of scale on our measures, the principal curvatures were first normalized using the volume, V,

of the convex hull of the binary WM mask. The curvatures were multiplied by radius, r, of a sphere with equivalent volume i.e. $r = \sqrt[3]{3V/4\pi}$, which has the desirable property of producing a dimensionless quantity which is 1 for any point on the surface of a sphere.

Three fundamental curvature-based shape descriptors were evaluated at each surface voxel: mean curvature (H), which intuitively expresses convexity and concavity around a point as positive and negative values; Gaussian curvature (K), which expresses elliptical and hyperbolic points as positive and negative values; and curvedness (C), a positive scalar that is related to the bending energy of a surface when deformed from a flat plate (van Vliet and Verbeeck, 1993). Cortical gyrification was then characterised by computing eight folding measures adapted from (Rodriguez-Carranza et al., 2008), each of which summarised a shape descriptor over the whole WM surface yielding a single scalar value (Table 1). These included global mean curvature (H_G), global Gaussian curvature (K_G), global curvedness (C_G), mean curvature L_2 norm (H_N), Gaussian curvature L_2 norm (K_N), intrinsic Gaussian curvature (K_I), mean curvature norm ratio (H_R) and Gaussian curvature norm ratio (K_R). Each of these measures was formulated to ensure its independence of the surface area. H_G expresses the predominant convexity or concavity of a surface, K_G expresses whether the surface is chiefly elliptical or hyperbolic, while C_G gives the average bending energy over the surface. Curvature norms, K_N and H_N, describe the overall degree of gyrification over the surface. Intrinsic Gaussian curvature, K_I, quantifies the degree of folding in only elliptical regions. Finally, curvature norm ratios, H_R and K_R, also quantify global gyrification and are intrinsically independent of surface area.

Local folding patterns were also investigated by applying these measures in localised regions. An atlas with 9 cortical labels derived from (Gousias et al., 2012) was registered to each brain, parcellating the WM surface into different anatomical regions (Figure 7). Folding measures were then computed over each region separately.

A voxel weighting term, w, was introduced to address the issue of unequal sampling density over the surface due to its voxel-based representation. For example, a plane perpendicular to an axis is more densely sampled than a diagonal plane. To compensate for this, we use the method proposed by (Windreich et al., 2003) which was used to weight voxels based on their connectivity configuration with background and other surface voxels.

Shape Descriptors	
Mean curvature	$H = (k_1 + k_2)/2$
Gaussian curvature	$K = k_1 k_2$
Curvedness	$C = \sqrt{(k_1^2 + k_2^2)/2}$

Folding Measures	
Global mean curvature	$H_G = \sum_{x_i} Hw \,/\, \sum_{x_i} w$
Global Gaussian curvature	$K_G = \sum_{x_i} Kw \,/\, \sum_{x_i} w$
Global curvedness	$C_G = \sum_{x_i} Cw \,/\, \sum_{x_i} w$
Mean curvature L_2 norm	$H_N = \sqrt{\sum_{x_i} H^2 w \,/\, \sum_{x_i} w}$
Gaussian curvature L_2 norm	$K_N = \sqrt[4]{\sum_{x_i} K^2 w \,/\, \sum_{x_i} w}$
Intrinsic Gaussian curvature	$K_I = \sqrt{\sum_{x_i \in K^+} Kw \,/\, \sum_{x_i \in K^+} w}$
Mean curvature norm ratio	$H_R = \sum_{x_i} H^2 w \,/\, \sum_{x_i} \lvert H \rvert w$
Gaussian curvature norm ratio	$K_R = \sqrt{\sum_{x_i} K^2 w \,/\, \sum_{x_i} \lvert K \rvert w}$

Table 1: **Summary of curvature-based shape descriptors and folding measures.** Three curvature-based shape descriptors were evaluated at each surface voxel, Gaussian curvature, mean curvature and curvedness. The degree of gyrification was quantified using eight folding measures, each of which were formulated by summarising a shape descriptor over all surface voxels, x_i. A weighting, w, was applied to each voxel which depended on its surface voxel configuration. Each of these measures was formulated to ensure its independence of the surface area. Note, K^+, denotes the set of voxels with positive Gaussian curvature.

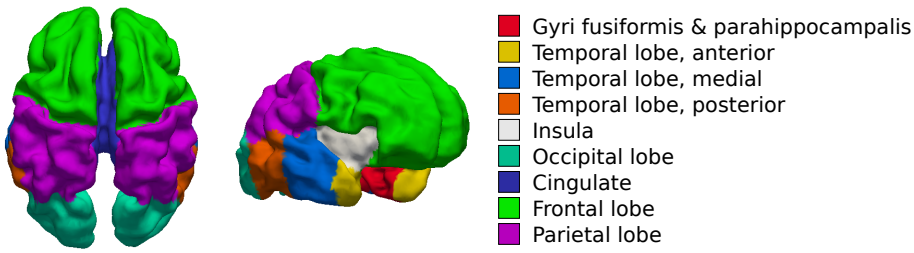


Figure 7: **Regional parcellation.** Example of a brain parcellated into nine different regions by registering an anatomical atlas.

3. Results & Discussion

3.1. Segmentation

To assess the accuracy of the automated segmentation procedure, 16 of the 80 subjects were manually segmented by a clinical expert and the discrepancy of the GM/WM boundary was measured. Details of the segmentation procedure can be found in (Kyriakopoulou et al., In Press). These subjects were chosen at evenly spaced intervals throughout the cohort age range. A surface mesh was extracted from the manual segmentation using the marching cubes algorithm (Lorensen and Cline, 1987) and a boundary distance map was extracted from the automated segmentation and interpolated to give a local distance at each vertex on the mesh. The mean distance error was then computed over all mesh vertices for each of the sixteen subjects (Table 2). The automated segmentation mean error was approximately 0.8mm - 1mm. Intra and inter-rater error was assessed and was found to be of the order of 0.13mm and 0.16mm respectively. There was a systematic bias towards a thicker estimate of the cortex in the automated segmentations, but despite this the surface meshes based on automated segmentation appeared to be a good approximation of the manual based mesh (Figure 8).

GA [weeks]	Error [mm]	Standard Deviation [mm]	95th percentile [mm]
22.4	0.74	0.64	2.11
23.6	0.90	0.69	2.24
24.0	0.68	0.55	1.84
24.3	0.70	0.55	1.81
26.7	0.79	0.76	2.31
26.9	0.75	0.57	1.75
27.6	0.95	0.62	2.16
28.1	0.73	0.66	1.85
28.1	0.92	0.75	2.41
29.9	1.08	0.74	2.5
30.4	0.88	0.73	2.25
31.3	0.94	0.88	2.97
31.6	0.95	0.72	2.13
32	1.11	0.96	3.09
34.1	1.07	0.96	2.91
36.4	0.85	0.84	2.50

Table 2: **Mean segmentation error**. The distance errors of the automated segmentations with respect to the manually generated ground truth segmentation for sixteen subjects.

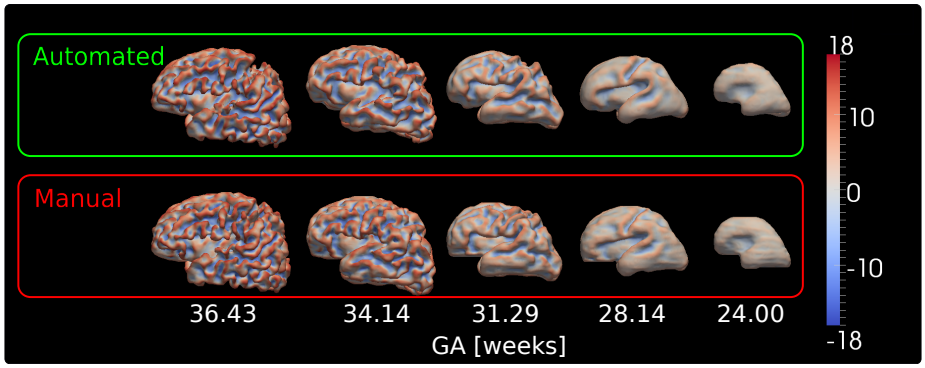


Figure 8: **Comparison of cortical surface meshes for manual and automatic segmentations.** Mesh representations of the surfaces were construction using the marching cubes algorithm for visualization purposes. Colour mapping depicts normalized mean curvature (H) at each point on the cortical surface.

3.2. Folding Measures

For each of the 80 fetuses, eight folding measures were computed over the automatically delineated WM/GM boundary. The folding measures based on auto-

matic segmentation replicated growth trends based on manual segmentations and were also self-consistent (Figure 9). However, the folding measures based on automatic segmentation consistently underestimated the values of folding measures derived from manual segmentation. Furthermore, the discrepancy increased with GA and this is at least partly due to the systematic segmentation bias that led to an underestimation of the WM volume, which in turn increases the magnitude of the normalized principal curvatures that are the basis of our folding measures. This could also be explained by the generally lower segmentation accuracy at older GAs as the cortex becomes more convoluted and the cortical boundary becomes less distinct with increasing PV. This is likely to decrease the complexity of the estimated surface and reduce the estimated volume, further decreasing folding measures. The reduction in growth rate in the latter stages of development is therefore likely to be overestimated and folding measures may converge to a higher value.

Folding measures that reflect the degree of gyrification (C_G, H_N, K_N, K_I, H_R, K_R) showed a strong positive correlation with GA (Figure 9). This suggests that these measures are effective for quantifying cortical folding and neurological development. Moreover, the correlation between these folding measures and GA is stronger than that of GA and volume (Table 3a), showing that neural development is more strongly linked to cortical complexity than brain size in this age range. These measures suggest that folding initially proceeds slowly, followed by a period of rapid growth from 25 weeks onwards, which begins to slow towards birth. This reflects what has been observed visually (Garel et al., 2001), the brain is initially lissencephalic at 22 weeks with the sylvian fissure present as a shallow depression. Then primary sulci, such as the central sulcus appear first around 24-25 weeks before lobar gyrification intensifies around 30 weeks. By 34 weeks GA, all of the primary and most of the secondary sulci are present.

This type of asymptotic growth, where the rate is slowest at the start and end of a time period is accurately modelled by a Gompertz function, which was first proposed by Gompertz in 1825. This function has been successfully used to model various phenomena that reach a saturation point, such as mobile phone users (Tao, 2010) and tumour cells (Laird, 1964). In the instance of tumour growth, a cellular population initially increases at an exponential rate but growth slows as nutritional supplies deplete. Gyrification is also a growth process that must eventually reach a saturation point. In this case, folding is limited by the thickness of the cortical sheet, which cannot be folded indefinitely due to physiological limitations. A Gompertz function is, therefore, a good candidate for modelling the rate of gyrification through gestation.

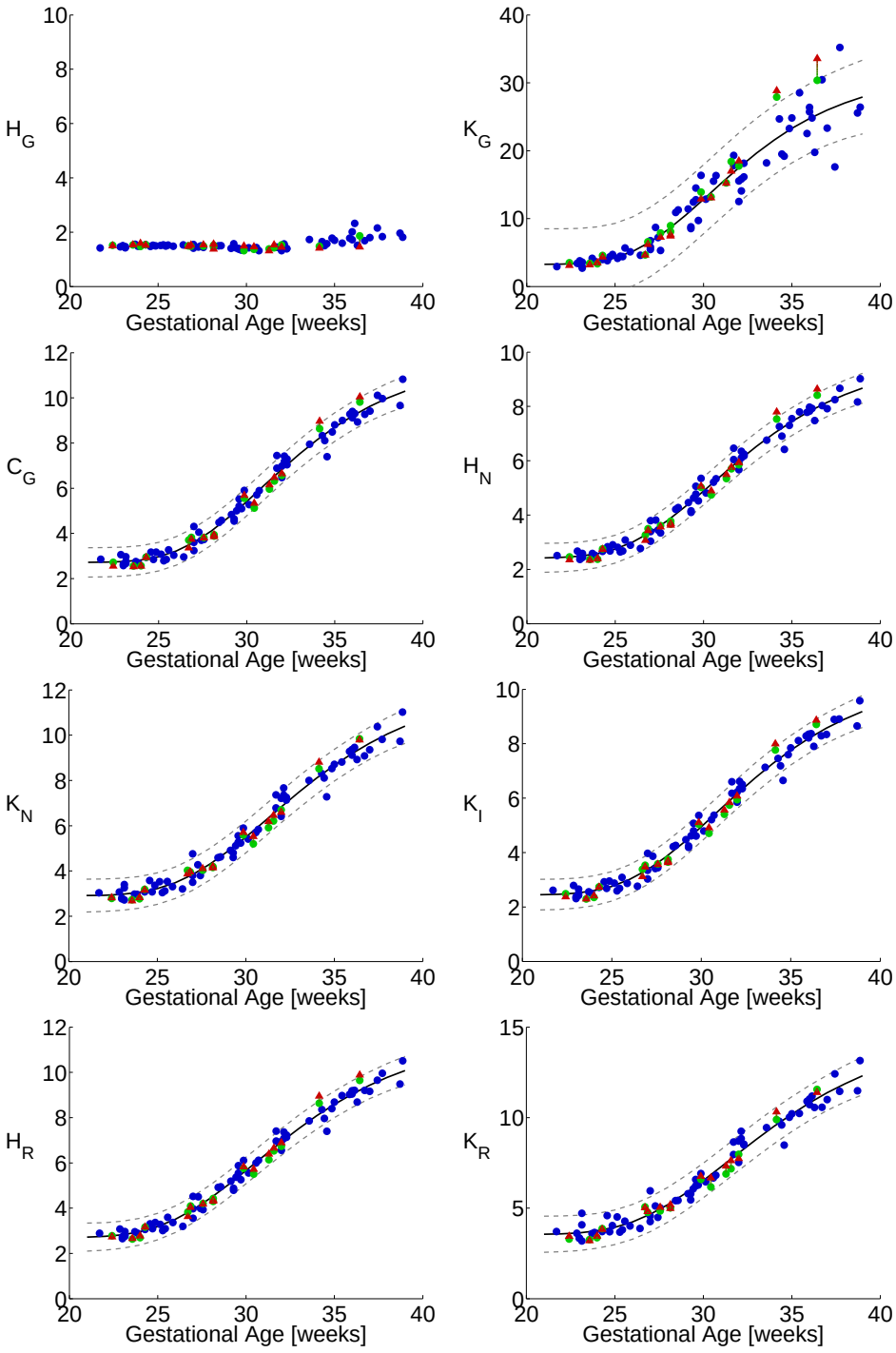


Figure 9: **Global folding measures with gestational age.** Curvature values computed from automated segmentations are shown as either green or blue circles, whereas curvatures computed from manual segmentations are shown as red triangles. For each subject that was segmented both manually and automatically, the corresponding points are shown in red and green respectively and joined together by a line. The remaining subjects that were only segmented automatically are represented by a blue circle. A Gompertz function (black line) was fitted to each plot except global mean curvature, which did not exhibit Gompertz like growth. Upper and lower confidence intervals are shown with dashed grey lines.

3.3. Gompertz Model

We examine the suitability of the Gompertz-like function (Equation 1) for modelling cortical gyrification and predict the value of a folding measure $y = f(t)$ given a gestational age t with the non-negative parameters β_i. Parameter β_1 is the initial value of y when the brain is lissencephalic before folding commences, $\beta_1 + \beta_2$ is the upper limit of y when folding has ceased increasing, β_3 controls the growth rate that determines the increase of y during gyrification and β_4 is a centring parameter, which determines the gestational age where peak growth occurs. The fitting of the model and estimation of the confidence intervals was carried out using the Statistics Toolbox functions nlinfit and nlpredci in the MatLab Software package (MATLAB and Toolbox, 2011) .

$$\hat{t} = t - \beta_4.$$

$$f(\hat{t}) = \beta_1 + \beta_2 e^{-e^{-\beta_3 \hat{t}}}. \tag{1}$$

$$f'(\hat{t}) = \beta_2 \beta_3 e^{-\beta_3 \hat{t} - e^{-\beta_3 \hat{t}}}. \tag{2}$$

To determine whether a Gompertz function accurately models the relationship between folding measures and gestational age, the goodness of fit for unseen data was evaluated by performing 100 iterations of 10 fold cross validation. The data were partitioned into 10 equal folds and each fold was held back in turn while a model was fitted to the remaining data and a residual was predicted for each omitted observation. The sum of square prediction residuals, $\hat{\epsilon}$, gave an evaluation measure, which was averaged over 100 iterations. The Gompertz model was compared against two other models: a linear function and a quadratic function (Table 3), to determine whether a simpler model could fit the data as well. Globally, Gompertz functions were found to fit the data closely ($R^2 = 0.99$) and considerably better than a simple quadratic or linear fit (Table 3a). On average, adopting a Gompertz model reduced $\hat{\epsilon}$ by a third over a quadratic model and almost a half over a linear model.

It must be noted that while gyrification must cease increasing at some point, and the relationship between folding measures and age is expected to be well modelled by a Gompertz function, systematic segmentation errors towards the latter gestational ages, where segmentation is particularly challenging, may lead to a false model of growth. If the segmentation algorithm fails to segment an increasingly complex cortical surface, the curvature measures may be underestimated and an asymptotic value may be reached before gyrification has actually ceased, and the value of model parameter β_2 will be under-estimated.

Measure	Gompertz		Quadratic		Linear	
	$\hat{\epsilon}$	R^2	$\hat{\epsilon}$	R^2	$\hat{\epsilon}$	R^2
C_G	**8.4**	**0.991**	16.3	0.983	22.7	0.977
H_N	**5.5**	**0.992**	10.2	0.984	11.9	0.982
K_N	**10.3**	**0.989**	15.4	0.983	22.6	0.975
K_I	**6.1**	**0.992**	11.3	0.985	15.0	0.980
H_R	**7.2**	**0.992**	12.3	0.986	14.3	0.984
K_R	**19.5**	**0.984**	24.8	0.980	35.3	0.971
Volume	-	-	-	-	-	0.958

(a) Model fit with different folding measures and volume

Region	Gompertz		Quadratic		Linear	
	$\hat{\epsilon}$	R^2	$\hat{\epsilon}$	R^2	$\hat{\epsilon}$	R^2
Global	**5.5**	**0.992**	10.2	0.984	11.9	0.982
Gyri fusiformis & parahippocampalis	**9.9**	**0.986**	15.8	0.978	20.7	0.971
Temporal lobe, anterior	23.5	0.950	22.2	0.953	**22.1**	**0.953**
Temporal lobe, medial	6.3	0.989	9.3	0.984	13.8	0.976
Insula	8.8	0.969	8.8	0.969	**9.3**	**0.967**
Occipital lobe	**6.3**	**0.990**	11.1	0.983	16.2	0.975
Temporal lobe, posterior	**10.5**	**0.988**	22.8	0.974	25.4	0.971
Cingulate	**19.0**	**0.959**	20.6	0.955	26.8	0.941
Frontal lobe	**5.8**	**0.991**	10.1	0.984	11.5	0.982
Parietal lobe	**7.3**	**0.991**	15.6	0.981	15.3	0.981

(b) Model fit in individual regions for H_N

Table 3: **Modelling the relationship of folding measures and GA.** Three models were fitted to the data: a Gompertz model, a quadratic model and a linear model. The predicted sum of square residuals $\hat{\epsilon}$ and coefficient of determination R^2 were estimated to evaluate the fit of each model.

Given the precise timing of sulcal development(Garel et al., 2001), a high correlation between age and folding measures is expected . However, we rule out the possibility of a strong bias towards the atlas priors further reducing variability and falsely strengthening the observed correlation. Each atlas frame is constructed by averaging FFDs for contributing subjects towards other subjects within a time window. The anatomy and segmentations are then transformed to an average atlas space using this deformation. The resulting atlas template does not exhibit the typical anatomy of each subject but rather an average, with reduced variability of gyral location and an overall smoother cortical surface. Therefore, curvature measures computed from atlas priors are expected to be lower than from an aged-matched subject. If the atlas priors were to dominate the segmentation and the algorithm failed to capture the subject's anatomy, then this would be reflected in the curvature measures. Figure 10 shows that this is not the case, as the curvatures computed from atlas priors are considerably lower than those computed for individual subjects with a similar age.

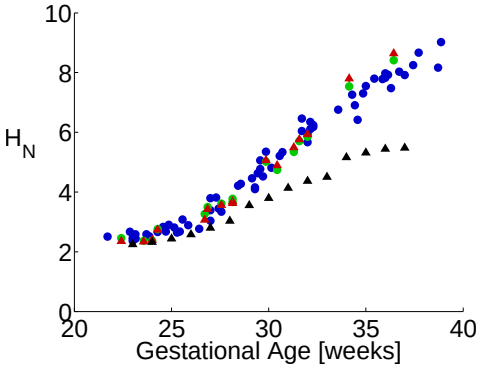
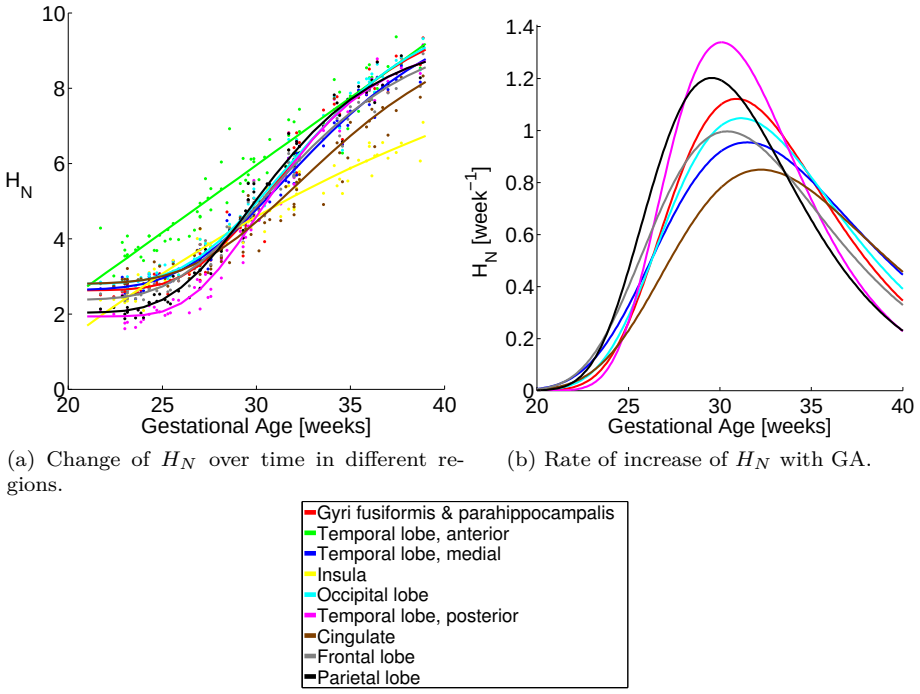


Figure 10: **Folding measures computed from atlas priors.** The mean curvature L_2 norm (H_N) is shown for each subject computed from automated segmentations (either green or blue circles), manual segmentations (red triangles) and atlas priors (black triangles). For subjects segmented both manually and automatically, the corresponding points are shown in red and green respectively and joined together by a line. Note the values of H_N computed from atlas priors are much lower than for subjects of a similar age.

3.4. Regional Folding

Examining folding measures locally revealed regional differences in sulcation. Lobar regions exhibited a similar growth pattern to what was observed globally and were also well modelled by a Gompertz function (Figure 11a). The insula and the anterior temporal lobe, however, did not display the same behaviour. A steady increase in folding measures with GA was observed in the anterior temporal lobe over the the time period, whereas the insula remained relatively smooth with folding measures not increasing with GA at the same rate as other regions. Consequently, a linear model gave a better fit for these regions. (Table 3b).

To examine the rate of growth across gestational ages we evaluated the derivative of the fitted Gompertz functions (Equation 2) for H_N in lobar regions over the observed time period. Peak growth, which occurred at $t = \beta_4$, was then estimated for each model. The rate of growth peaked around 30 weeks gestational age for all lobar regions, suggesting that this is a critical time period and any developmental disturbances could have a greater impact around this point (Figure 11b). Regional differences in growth rates were also observed, with the posterior temporal lobe and parietal lobe showing the highest peak growth, with other regions such as the frontal lobe and medial temporal lobe developing more slowly, and the cingulate considerably slower (Table 4).



(a) Change of H_N over time in different regions.

(b) Rate of increase of H_N with GA.

Figure 11: **Regional differences in folding measures.** (a) The change in mean curvature L_2 norm (H_N) over individual regions with gestational age is shown. All lobar regions expect the insula and anterior temporal lobe exhibit a Gompertz like growth pattern. (b) The rate of growth across gestational ages was computed from the derivative of the fitted Gompertz functions for H_N in lobar regions. The rate of growth peaked around 30 weeks gestational age for all regions, however there are differences in growth rates.

Region	Peak Growth	GA
Temporal lobe, posterior	1.00	29.8
Parietal lobe	0.90	29.6
Gyri fusiformis & parahippocampalis	0.84	30.9
Occipital lobe	0.78	31.2
Frontal lobe	0.74	30.4
Temporal lobe, medial	0.71	31.5
Cingulate	0.63	32.2

Table 4: **Relative peak growth rate in lobar regions.** The peak increase of H_N is shown, relative to the posterior temporal lobe. Peak growth occurred around 30 weeks gestational age for all lobar regions and regional differences in growth rates are apparent.

3.5. Physiological Age

The strong relationship between GA and folding measures allowed an accurate prediction of GA to be made based on an observed folding measure. The inverse of the Gompertz function fitted for measure H_N was taken to predict GA from the value of H_N observed for each fetus. Prediction accuracy was inherently lower where the rate of folding was slowest. When gyrification changes very little, folding measures become a poor indicator of GA, as fetuses in close proximity have a similar degree of gyrification. Therefore we expect age prediction to be poor towards the extremities of the age range in our cohort (Figure 12). The mean prediction error was 0.47 weeks ($\pm$ 0.40) in the age range 24-37 weeks and 0.88 weeks ($\pm$ 0.61) outside of this range, where the rate of folding was considerably slower. Given that the real GA is unknown and is estimated from either an obstetric ultrasound scan or the last menstrual period, the measured prediction accuracy of the classifier is limited and may not reflect the real accuracy.

Predicting the real GA of a fetus from an MR image is not a genuinely useful application of this work, given that GA may be estimated by obstetric ultrasound. However, by estimating the age of a fetus based on the observed anatomy, we are in effect estimating the physiological age i.e. the average age of a normal fetus with the equivalent level of neurological development. A difference in real GA and estimated GA could, therefore, indicate a delay or acceleration from normal development, associated with some pathology.

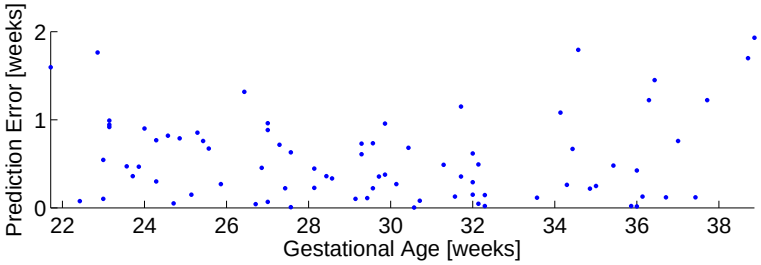


Figure 12: **GA prediction error.** Gestational age was predicted from the observed value of H_N by taking the inverse of the Gompertz function fitted to the data for H_N.

4. Conclusions

In this paper, we have presented a complete framework to quantify folding patterns in the fetus from *in utero* MRI. Using a robust slice-to-volume reconstruction technique (Xue et al., 2007) we obtained coherent 3D volumes of 80 fetuses over a wide gestational age range (22-39 weeks) and accurately delineated intra-cranial matter using the patch-based method of (Eskildsen et al.,

2012). To estimate the boundary between WM and cortical GM, we first applied an EM-MRF segmentation technique (Ledig et al., 2012), which accounted for mislabelled sulcal PV through a second order MRF regularization, to accurately segment the fetal brains into seven fundamental structures. Sixteen of the 80 fetal brain images were segmented manually to validate the automated procedure and the average discrepancy between the two cortical surfaces was found to be around 1 mm. For each fetus, principal curvatures were computed over an estimated surface using the structure tensor and Hessian matrix (Rodriguez-Carranza et al., 2008). Three curvature-based shape descriptors, derived from the principal curvatures, were then evaluated at each surface voxel. Eight folding measures were computed for each fetus, each of which summarised a shape descriptor over the surface to produce a single scalar value.

We revealed a strong positive correlation between several cortical folding measures and GA. Moreover, these measures were better correlated with GA than volume. This suggests that neurological development is more closely linked with the convolution of the cortex than the size of the cerebrum. The increase in folding measures with GA reflects the rate of gyrification that has been observed visually (Garel et al., 2001), with gyrification accelerating rapidly between 25 and 30 weeks. A Gompertz function is an ideal candidate for modelling the relationship between folding measures and GA, and we observed a better fit compared to a simple linear or quadratic model. However, we also observed that the values of curvature measures were systematically underestimated at older gestational ages by a small margin when using automated methods. Therefore, it is likely that curvature measures will reach an asymptotic level while gyrification is continuing. Thus, further work is needed to assess the reliability of predictions based on this model at times when the gyrification process is nearing completion.

Employing a Gompertz model allowed an accurate prediction of GA of a fetus with a mean error of approximately half a week between 24-37 weeks GA. The neurological development of a fetus can then be characterised using the notion of physiological age and a developmental delay or acceleration can be quantified in weeks. The model also allowed us to analyse growth rates over time, and peak gyrification was observed in all lobar regions around 30 weeks. Therefore this period may be a critical time for neurological development.

By characterising healthy folding patterns in the fetus, we can begin to uncover regions of abnormal folding that may be associated with various neurological disorders such as autism. This could be particularly beneficial in aiding diagnosis where subtle abnormalities may be overlooked during reporting, or where a 2D in-plane analysis makes assessment difficult. Automated detection of anomalies could lead to earlier diagnosis which will lead to better care for the patient. Furthermore, by establishing normal folding patterns in the fetus we can identify abnormal folding patterns in preterm neonates who have a higher risk of functional disabilities such as cognitive impairment (Johnson et al., 2009).

Appendix A. Patch-based Brain Extraction

To label an image voxel x_i we took a cubic patch P_i centred at x_i and computed the similarity to all patches $P_{a,j}$ of a single atlas a, in a local cubic neighbourhood η_i of x_i. Firstly, patch similarity was compared using the structural similarity (SS) metric (Wang et al., 2004): where μ is the patch mean and σ is the patch standard deviation (Equation A.1). Atlas patches with a SS $\leq$ 0.95 were discarded.

$$SS = \frac{2\mu_i\mu_{a,j}}{\mu_i^2 + \mu_{a,j}^2} \times \frac{2\sigma_i\sigma_{a,j}}{\sigma_i^2 + \sigma_{a,j}^2}. \tag{A.1}$$

A vote was taken from the remaining atlas patch labels $l(x_{a,j})$ weighted by their similarity to the target patch, $w(P_i, P_{a,j})$, to obtain a label $L_{a,i}$ for atlas a:

$$L_{a,i} = \left\lfloor 0.5 + \frac{\sum_j w(P_i, P_{a,j}) l(x_{a,j})}{\sum_j w(P_i, P_{a,j})} \right\rfloor. \tag{A.2}$$

$$w(P_i, P_{a,j}) = \exp(\frac{-SSD(P_i, P_{a,j})}{h_a}). \qquad h_a = min(SSD(P_i, P_{a,j})). \tag{A.3}$$

Finally, after computing a binary labelling $L_{a,i}$ using each atlas in turn, we then took a majority vote between all atlases to reach a final binary labelling $\hat{L}_i$:

$$\hat{L}_i = \left\lfloor 0.5 + \frac{1}{N} \sum_a L_{a,i} \right\rfloor. \tag{A.4}$$

The limited variation in the shape of the cranial cavity was exploited to speed up extraction by defining a region of interest (ROI) in which to label voxels. The union of all N atlas labels a_i defined a region where brain voxels were possible, $a_1 \cup a_2 \cup \ldots \cup a_N$, while the intersection defined a region where brain voxels were almost certain, $a_1 \cap a_2 \cap \ldots \cap a_N$. We took the set difference,$(a_1 \cup a_2 \cup \ldots \cup a_N) \backslash (a_1 \cap a_2 \cap \ldots \cap a_N)$, to give the ROI over which to apply patch-based brain extraction and dilated it to account for possible variation not captured by the atlas set. The most similar atlases were chosen to label a new image based on their global similarity by taking the sum of squared differences in this ROI.

Appendix B. Principal Curvatures

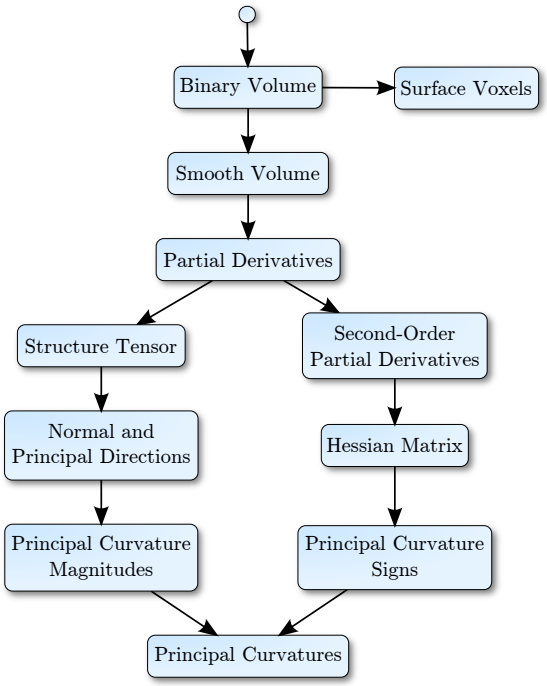


Figure B.13: **Calculating the principal curvatures.** The surface voxels, where the principal curvatures are evaluated, are defined as any foreground voxel that shares a face with a background voxel within the binary WM volume. The binary volume was then smoothed to produce a continuous scalar field on which to calculate the image derivatives. The calculation of the principal curvature magnitudes and signs is then split into two processes. The structure tensor is computed from the image derivatives and allows the principal curvature directions and magnitudes to be computed. The signs of the principal curvatures are then recovered using the Hessian matrix.

For each surface voxel at location x_i, the principal curvatures k_1 and k_2 were calculated using the structure tensor and Hessian matrix (see Figure B.13 for an overview). The structure tensor is a matrix derived from the partial derivatives of a scalar function and can be used to analyse the gradient orientation of a neighbourhood. We obtain a scalar field $I \in \mathbb{R}$ from minimal smoothing of

the binary WM volume to produce a continuous representation and define the partial derivative vector $\mathbf{v} = \nabla I$.

The structure tensor $\mathbf{S}(x_i)$ for a surface voxel x_i is defined as the sum of outer products $\mathbf{S}^0(x_j)$ of the derivative vector $\mathbf{v}(x_j)$ with Gaussian weighting $w(x_i, x_j)$ for each voxel x_j in a local neighbourhood η_i:

$$\mathbf{S}(x_i) = \sum_{x_j \in \eta_i} \mathbf{S}^0(x_j)\, w(x_i, x_j).$$

$$\mathbf{S}^0(x_j) = \mathbf{v}(x_j)\, \mathbf{v}(x_j)^T.$$

$$\mathbf{v}(x_j) = \left[\frac{\partial I}{\partial x}(x_j)\ \frac{\partial I}{\partial y}(x_j)\ \frac{\partial I}{\partial z}(x_j) \right]^T.$$

Eigen-decomposition of the tensor summarises the gradient directions. Assuming an ordering of the eigenvalues: $\lambda_1 > \lambda_2 > \lambda_3$, the gradient, the eigenvector $\mathbf{e}_1$ and $\mathbf{n}$ the surface normal are all aligned. The eigenvectors $\mathbf{e}_2$ and $\mathbf{e}_3$ are aligned with the maximum and minimum curvature directions $\mathbf{t}_1$ and $\mathbf{t}_2$ respectively.

Unfortunately, we cannot differentiate the normal with respect to the principal directions to get the principal curvatures as the vector field obtained from the first eigenvectors is discontinuous i.e. $\mathbf{e}_1 = \pm\mathbf{n}$. Instead, a mapping $\mathbf{M}(\mathbf{e}_1) = \mathbf{e}_1\mathbf{e}_1^T / \|\mathbf{e}_1\|$ is used to calculate the magnitudes and the sign of the curvature is recovered using the Hessian matrix. The norm of the derivative is linearly related to the norm of the mapped eigenvector $\mathbf{M}(\mathbf{e}_1)$ by a constant, $\sqrt{2}$. We differentiate each element m_{ij} of the matrix $\mathbf{M}$ with respect to the three coordinate axes ϕ^1, ϕ^2 and ϕ^3:

$$|k_{1,2}| = \frac{1}{\sqrt{2}} \left\| \sum_{k=1}^{3} \frac{\partial m_{ij}}{\partial \phi^k} e_{2,3}^k \right\|.$$

$$\|\mathbf{M}\| := \text{Frobenius norm of } \mathbf{M}.$$

To recover the sign of a principal curvature $k_{1,2}$ we employ the Hessian matrix $\mathcal{H}$ (Equation B.1) and examine the second derivative along the corresponding principal direction $\mathbf{t}_{1,2}$, specifically we evaluate $\mathbf{t}^T\mathcal{H}\mathbf{t}$. If the result is less than zero, the principal curvature $k_{1,2}$ is positive (i.e. the surface is convex), otherwise the curvature is negative (surface is concave).

$$\mathcal{H} = \begin{bmatrix} \frac{\partial I}{\partial x^2} & \frac{\partial I}{\partial xy} & \frac{\partial I}{\partial xz} \\ \frac{\partial I}{\partial xy} & \frac{\partial I}{\partial y^2} & \frac{\partial I}{\partial yz} \\ \frac{\partial I}{\partial xz} & \frac{\partial I}{\partial yz} & \frac{\partial I}{\partial z^2} \end{bmatrix}. \tag{B.1}$$

Acknowledgements

This research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

B. Caldairou, N. Passat, P. Habas, C. Studholme, M. Koob, J.L. Dietemann, and F. Rousseau. Segmentation of the cortex in fetal MRI using a topological model. In *IEEE International Symposium on Biomedical Imaging: ISBI*, pages 2045–2048, 2011. doi: 10.1109/ISBI.2011.5872814.

G. Egaña-Ugrinovic, M. Sanz-Cortes, F. Figueras, N. Bargalló, and E. Gratacós. Differences in cortical development assessed by fetal MRI in late-onset intrauterine growth restriction. *American Journal of Obstetrics and Gynecology*, In Press. doi: 10.1016/j.ajog.2013.04.008.

S.F. Eskildsen, P. Coupé, V. Fonov, J.V. Manjón, K.K. Leung, N. Guizard, S.N. Wassef, L.R. Østergaard, and D.L. Collins. BEaST: Brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373, 2012. doi: 10.1016/j.neuroimage.2011.09.012.

C. Garel, E. Chantrel, H. Brisse, M. Elmaleh, D. Luton, J. Oury, G. Sebag, and M. Hassan. Fetal cerebral cortex: Normal gestational landmarks identified using prenatal mr imaging. *American Journal of Neuroradiology*, 22(1):184–189, 2001.

S. Ghai, K.W. Fong, A. Toi, D. Chitayat, S. Pantazi, and S. Blaser. Prenatal US and MR imaging findings of lissencephaly: Review of fetal cerebral sulcal development. *Radiographics*, 26(2):389–405, 2006. doi: 10.1148/rg.262055059.

A. Gholipour, J.A. Estroff, and S.K. Warfield. Robust super-resolution volume reconstruction from slice acquisitions: Application to fetal brain MRI. *IEEE Transactions on Medical Imaging*, 29(10):1739–1758, 2010. doi: 10.1109/tmi.2010.2051680.

A. Gholipour, A. Akhondi-Asl, J.A. Estroff, and S.K. Warfield. Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly. *NeuroImage*, 60(3):1819–1831, 2012. doi: 10.1016/j.neuroimage.2012.01.128.

C.M. Glastonbury and A.M. Kennedy. Ultrafast MRI of the fetus. *Australasian Radiology*, 46(1):22–32, 2002. doi: 10.1046/j.1440-1673.2001.00990.x.

I.S. Gousias, A.D. Edwards, M.A. Rutherford, S.J. Counsell, J.V. Hajnal, D. Rueckert, and A. Hammers. Magnetic resonance imaging of the newborn brain: Manual segmentation of labelled atlases in term-born and preterm infants. *NeuroImage*, 62(3):1499–1509, 2012. doi: 10.1016/j.neuroimage.2012.05.083.

P.A. Habas, K. Kim, F. Rousseau, O.A. Glenn, A.J Barkovich, and C. Studholme. Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses. *Human Brain Mapping*, 31(9): 1348–1358, 2010. doi: 10.1002/hbm.20935.

P.A. Habas, J.A. Scott, A. Roosta, V. Rajagopalan, K. Kim, F. Rousseau, A.J. Barkovich, O.A. Glenn, and C. Studholme. Early folding patterns and asymmetries of the normal human brain detected from in utero MRI. *Cerebral Cortex*, 22(1):13–25, 2012. doi: 10.1093/cercor/bhr053.

J.W. Hand, Y. Li, and J.V. Hajnal. Numerical study of RF exposure and the resulting temperature rise in the foetus during a magnetic resonance procedure. *Physics in Medicine and Biology*, 55(4):913–930, 2010. doi: 10.1088/0031-9155/55/4/001.

I.A. Hosny and H.S. Elghawabi. Ultrafast MRI of the fetus: an increasingly important tool in prenatal diagnosis of congenital anomalies. *Magnetic Resonance Imaging*, 28(10):1431–1439, 2010. doi: 10.1016/j.mri.2010.06.024.

F.D. Jacob, P.A. Habas, K. Kim, J. Corbett-Detig, D. Xu, C. Studholme, and O.A. Glenn. Fetal hippocampal development: analysis by magnetic resonance imaging volumetry. *Pediatric Research*, 69(5 pt 1):425–429, 2011. doi: 10.1203/pdr.0b013e318211dd7f.

S. Jiang, H. Xue, A. Glover, M. Rutherford, D. Rueckert, and J.V. Hajnal. MRI of moving subjects using multislice snapshot images with volume reconstruction (SVR): Application to fetal, neonatal, and adult brain studies. *IEEE Transactions on Medical Imaging*, 26(7):967–980, 2007. doi: 10.1109/tmi.2007.895456.

S. Johnson, E. Hennessy, R. Smith, R. Trikic, D. Wolke, and N. Marlow. Academic attainment and special educational needs in extremely preterm children at 11 years of age: the EPICure study. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 94(4):F283–F289, 2009. doi: 10.1136/adc.2008.152793.

K. Kim, P.A. Habas, F. Rousseau, O.A. Glenn, A.J. Barkovich, and C. Studholme. Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation. *IEEE Transactions on Medical Imaging*, 29(1): 146–158, 2010. doi: 10.1109/tmi.2009.2030679.

M. Kuklisova-Murgasova, P. Aljabar, L. Srinivasan, S.J. Counsell, V. Doria, A. Serag, I.S. Gousias, J.P. Boardman, M.A. Rutherford, A.D Edwards, J.V. Hajnal, and D. Rueckert. A dynamic 4D probabilistic at-

las of the developing brain. *NeuroImage*, 54(4):2750–2763, 2011. doi: 10.1016/j.neuroimage.2010.10.019.

M. Kuklisova-Murgasova, G. Quaghebeur, M.A. Rutherford, J.V. Hajnal, and J.A. Schnabel. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Medical Image Analysis*, 16(8):1550–1564, 2012. doi: 10.1016/j.media.2012.07.004.

V. Kyriakopoulou, D. Vatansever, S. Elkommos, S. Dawson, A. McGuinness, J. Allsop, Z. Molnár, J. Hajnal, and M. Rutherford. Cortical overgrowth in fetuses with isolated ventriculomegaly. *Cerebral Cortex*, In Press. doi: 10.1093/cercor/bht062.

A.K. Laird. Dynamics of tumour growth. *British journal of cancer*, 18(3): 490–502, 1964.

C. Ledig, P. Aljabar, R. Wright, A. Serag, and D. Rueckert. Neonatal brain segmentation using second order neighborhood information. In *Workshop on Perinatal and Paediatric Imaging: PaPI, Medical Image Computing and Computer-Assisted Intervention: MICCAI*, pages 33–40, 2012.

W.E. Lorensen and H.E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *ACM Siggraph Computer Graphics*, volume 21, pages 163–169, 1987.

J.H. Lui, D.V. Hansen, and A.R. Kriegstein. Development and evolution of the human neocortex. *Cell*, 146(1):18–36, 2011. doi: 10.1016/j.cell.2011.06.030.

MATLAB and Statistics Toolbox. *version 7.13.0.564 (R2011b)*. The MathWorks Inc., Natick, Massachusetts, United States, 2011.

M. Murgasova, L. Dyet, D. Edwards, M.A. Rutherford, J.V. Hajnal, and D. Rueckert. Segmentation of brain MRI in young children. *Academic Radiology*, 14(11):1350–1366, 2007. doi: 10.1016/j.acra.2007.07.020.

L.G. Nyúl and J.K. Udupa. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine*, 42(6):1072–1081, 1999. doi: 10.1002/(sici)1522-2594(199912)42:6<1072::aid-mrm11>3.0.co;2-m.

D. Prayer, P.C. Brugger, and L. Prayer. Fetal MRI: techniques and protocols. *Pediatric Radiology*, 34(9):685–693, 2004. doi: 10.1007/s00247-004-1246-0.

V. Rajagopalan, J. Scott, P.A. Habas, K. Kim, J. Corbett-Detig, F. Rousseau, A.J. Barkovich, O.A. Glenn, and C. Studholme. Local tissue growth patterns underlying normal fetal human brain gyrification quantified in utero. *The Journal of Neuroscience*, 31(8):2878–2887, 2011. doi: 10.1523/jneurosci.5458-10.2011.

B. Rieger, F.J. Timmermans, L.J. van Vliet, and P.W. Verbeek. On curvature estimation of iso surfaces in 3D gray-value images and the computation of shape descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1088–1094, 2004. doi: 10.1109/tpami.2004.50.

C.E. Rodriguez-Carranza, P. Mukherjee, D. Vigneron, J. Barkovich, and C. Studholme. A framework for in vivo quantification of regional brain folding in premature neonates. *NeuroImage*, 41(2):462–478, 2008. doi: 10.1016/j.neuroimage.2008.01.008.

F. Rousseau, O.A. Glenn, B. Iordanova, C. Rodriguez-Carranza, D.B. Vigneron, J.A. Barkovich, and C. Studholme. Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images. *Academic Radiology*, 13(9):1072–1081, 2006. doi: 10.1016/j.acra.2006.05.003.

F. Rousseau, K. Kim, C. Studholme, M. Koob, and J.L. Dietemann. On super-resolution for fetal brain MRI. In *Medical Image Computing and Computer-Assisted Intervention: MICCAI*, volume 6362 of *Lecture Notes in Computer Science: LNCS*, pages 355–362, 2010. doi: 10.1007/978-3-642-15745-5_44.

D. Rueckert, L.I. Sonoda, C. Hayes, D.L. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999. doi: 10.1109/42.796284.

J. A. Scott, P. A. Habas, K. Kim, V. Rajagopalan, K.S. Hamzelou, J.M. Corbett-Detig, A.J. Barkovich, O.A. Glenn, and C. Studholme. Growth trajectories of the human fetal brain tissues estimated from 3D reconstructed in utero MRI. *International Journal of Developmental Neuroscience*, 29(5):529–536, 2011. doi: 10.1016/j.ijdevneu.2011.04.001.

J.A. Scott, P.A. Habas, V. Rajagopalan, K. Kim, A.J. Barkovich, O.A. Glenn, and C. Studholme. Volumetric and surface-based 3D MRI analyses of fetal isolated mild ventriculomegaly. *Brain Structure and Function*, 218(3):645–655, 2013. doi: 10.1007/s00429-012-0418-1.

A. Serag, V. Kyriakopoulou, P. Aljabar, S.J. Counsell, J.P. Boardman, M.A. Rutherford, A.D. Edwards, J.V. Hajnal, and D. Rueckert. A multi-channel 4D probabilistic atlas of the developing brain: Application to fetuses and neonates. In *Annals of the British Machine Vision Association*, number 3, pages 1–14, 2012.

S.M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, 2002. doi: 10.1002/hbm.10062.

J. Tao. Research on gompertz curve model used for mobile user growth. In *2010 International Conference on Educational and Network Technology*, pages 557–560, 2010. doi: 10.1109/ICENT.2010.5532100.

N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, and J.C. Gee. N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010. doi: 10.1109/tmi.2010.2046908.

K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, 1999. doi: 10.1109/42.811270.

L. J. van Vliet and P. W. Verbeeck. Curvature and bending energy in digitized 2D and 3D images. In *Scandinavian Conference on Image Analysis: SCIA*, volume 2, pages 1403–1410, 1993.

Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Medical Imaging*, 13(4):600–612, 2004. doi: 10.1109/tip.2003.819861.

R.C. Welsh, U. Nemec, and M.E. Thomason. Fetal magnetic resonance imaging at 3.0 T. *Topics in magnetic resonance imaging: TMRI*, 22(3):119–131, 2011. doi: 10.1097/rmr.0b013e318267f932.

J.P. De Wilde, A.W. Rivers, and D.L. Price. A review of the current use of magnetic resonance imaging in pregnancy and safety implications for the fetus. *Progress in Biophysics and Molecular Biology*, 87(2–3):335–353, 2005. doi: 10.1016/j.pbiomolbio.2004.08.010.

G. Windreich, N. Kiryati, and G. Lohmann. Voxel-based surface area estimation: from theory to practice. *Pattern Recognition*, 36(11):2531–2541, 2003. doi: 10.1016/S0031-3203(03)00173-0.

R. Wright, D. Vatansever, V. Kyriakopoulou, C. Ledig, R. Wolz, A. Serag, D. Rueckert, M.A Rutherford, J.V. Hajnal, and P. Aljabar. Age dependent fetal MR segmentation using manual and automated approaches. In *Workshop on Perinatal and Paediatric Imaging: PaPI, Medical Image Computing and Computer-Assisted Intervention: MICCAI*, pages 97–104, 2012.

H. Xue, L. Srinivasan, S. Jiang, M.A. Rutherford, A.D. Edwards, D. Rueckert, and J.V. Hajnal. Automatic segmentation and reconstruction of the cortex from neonatal MRI. *NeuroImage*, 38(3):461–477, 2007. doi: 10.1016/j.neuroimage.2007.07.030.